(54) Title: APPARATUS AND SYSTEM FOR CLASSIFYING AND CONTROL ACCESS TO INFORMATION
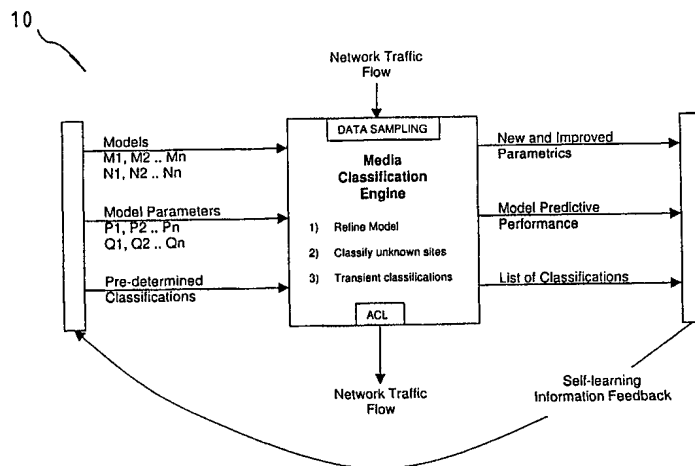
(57) Abstract

An apparatus (10) is provided for classifying information or content servers on a communications network including the Internet. The apparatus (10) comprises means for obtaining one or more transmission characteristics of information on a path of said communications network and analysing means for predicting a classification of said information based on said one or more transmission characteristics. Typically said one or more transmission characteristics include any one or more of network protocol, date and time stamps, size of transmission activities (text and image), content type of transmission activities, pattern seen within the content of the transmission and any other characteristic that can be employed for predicting classifications. The apparatus (10) can be adapted to classify user profiles in accordance with the predicted classification. A knowledge base of predetermined profiles can be included, and the analysing means is adapted to predict a classification based on a comparison between the profile of information to be classified and the predetermined profiles.

1

APPARATUS AND SYSTEM FOR CLASSIFYING AND
CONTROL ACCESS TO INFORMATION
TECHNICAL FIELD OF THE INVENTION

THIS INVENTION relates to apparatus and system for classifying information

5   on communications network and in particular but not limited to apparatus and

system for classifying content servers and for selectively controlling access to

classified content servers.

BACKGROUND OF THE INVENTION

The phenomenon growth of information technology has allowed many

10   people to have access to diverse information on communications networks. The

Internet in particular allows fetching of information from any cooperating

computers or content servers located in different parts of the world by simply

clicking references to the information. As the number of accessible computers or

content servers and the amount of information over the communications network

15   grow daily it becomes increasingly difficult to classify them manually.

Known systems for controlling the types of information accessible on a

network rely on comparing a requested destination with those on pre-determined

Access Control Lists (ACL) or on word matching to determine whether to allow or

deny access. This approach can be applied at the client node prior to requesting the

20   information or on any suitably intelligent network device capable of intercepting

the request or subsequent reply prior to it reaching the requester. For example, in

the case of an Internet browser running on a PC or work station, a request is made

for an Internet resource such as a web site. A software program for monitoring such

requests on the PC can be configured to scan a pre-determined list of site addresses

25   for a match. If found, access to the site may be denied and a suitable message is

then displayed informing the user that access is denied. Alternatively, the request

may be allowed to proceed, but as data are received from the site they are scanned

for checking a match with one or more sets of pre-determined words, word

fragments or phrases. If a match is found the site is not displayed on the computer

30   but instead there is shown a suitable message. Typically, this type of control

software is installed on a PC or work station which does not have particularly strict

2

access privileges. The control software can be easily removed, disabled or otherwise circumvented and thereby defeating the control system.

A network device capable of intercepting the request or reply to a request, such as a proxy server, may perform similar actions using the same methods of web site matching. This is usually maintained by a network administrator with strict access rights. Also, a network requiring clients to connect through the network device in order to access the network can have its content control enforced. This allows content control of multiple clients from one central point.

While these known systems do provide some access control abilities, there are several disadvantages. A system based on word or phrase matching can only match text and it therefore would allow access to undesired information comprising graphic images. Also, a single word may match a broad range of sites with quite different classes of information. As an example, when the word "sex" is used to match pornographic sites the system would also block access to other sites providing non offensive information such as articles on biology.

A system based on an access control list of prohibited sites is much more selective. Access can only be denied when attempting to access the sites which are included in the lists. While a suitably large list could bar access to a great deal of undesirable information it is difficult to keep up to date due to the rapid increase in the number of new sites and removal of sites.

The above systems also do not lend themselves to adaptation to other network protocols and services such as interactive chat, streaming video, email or encrypted data streams. Extending to different languages also poses a problem for globalisation of these systems.

OBJECT OF THE INVENTION

An object of the present invention is to alleviate or to reduce to a certain degree one or more of the above disadvantages.

Another object of the present invention is provide an apparatus/system for classifying user profiles.

3

SUMMARY OF THE INVENTION

In one aspect therefor the present invention resides in an apparatus for classifying information on communications network. The apparatus comprises means for obtaining one or more transmission characteristics of information on a

5   path of said communications network, and analysing means for predicting a classification of said information based on said one or more transmission characteristics.

In a second aspect therefor the present invention resides in an apparatus for classifying content servers which are accessible on a communications network. The

10   apparatus comprises means for obtaining one or more transmission characteristics of information provided by any of said content servers on a path of said communications network, and analysing means for predicting a classification of said information based on said one or more transmission characteristics.

In a third aspect therefor the present invention resides in a computer

15   program for classifying information which is accessible on a communications network. The program comprises means for obtaining one or more transmission characteristics of information on a path of said communications network, and analysing means for predicting a classification of said information based on said one or more transmission characteristics.

20   In a fourth aspect therefor the present invention resides in a computer program for classifying content servers which are accessible on a communications network. The apparatus comprises means for obtaining one or more transmission characteristics of information provided by any of said content servers on a path of said communications network, analysing means for predicting a classification of

25   said information based on said one or more transmission characteristics.

In a fifth aspect therefor the present invention resides in an apparatus/computer program for classifying user profiles of users accessing information or content servers on a communications network. The apparatus/computer program comprises means for obtaining one or more

30   transmission characteristics of information or information provided by any one of said content servers on a path of said communications network, analysing means

4

for predicting a classification of said information or said one content server based on said one or more transmission characteristics, and means for classifying user profile in accordance with the predicted classification.

The above invention may also comprise means for storing said one or more
5   transmission characteristics.

Typically said one or more transmission characteristics include any one or more of network protocol, date and time stamps, size of transmission activities (text and image), content type of transmission activities, pattern seen within the content of the transmission and any other characteristic that can be employed for predicting
10   classifications.

In preference said one or more transmission characteristics are obtained from network packets or fragments thereof.

It is also preferred that the analysing means includes profiling means for providing profiles of interactions based on said one or more transmission
15   characteristics. Typically said profiling means is arranged to process said one or more transmission characteristics for providing any one or more of frequency of interaction, duration of interaction, duration of absence of interaction, patterns of transmission, average number of http links within an object of related sites, average number of like sites visited within a time frame, and statistics from said other
20   characteristics, for forming interaction profiles. The analysing means can then use the profiles for predicting classifications.

The invention may have a knowledge base of predetermined profiles, and the analysing means is adapted to predict a classification based on a comparison between the profile of information to be classified and predetermined profiles.

25   Advantageously the invention may have means for updating the knowledge base so that the classification prediction may be enhanced following classifications.

In order that the present invention can be more readily understood and be put into practical effect reference will now be made to the accompanying drawings which illustrate one preferred embodiment of the invention and wherein:

30   BRIEF DESCRIPTION OF THE DRAWING

Figure 1 is a schematic diagram of the apparatus according to the invention;

5

Figure 2 is a table of selected data of captured packets of a search engine using the apparatus shown in Figure 1;

Figure 3 is a partial table of selected data of captured packets of a news web site using the apparatus shown in Figure 1;

5   Figure 4 is a table of selected data of captured packets of an entertainment web site using the apparatus shown in Figure 1;

Figure 5 is a table of selected data of captured packets of the web site of an e-commerce merchant using the apparatus shown in Figure 1;

Figure 6 is a table of selected data of captured packets of the web site of 10   another e-commerce merchant using the apparatus shown in Figure 1;

Figure 7 is a table of selected data of captured packets of a pornography web site using the apparatus shown in Figure 1;

Figure 8 is a table of selected data of captured packets of another pornography web site using the apparatus shown in Figure 1;

15   Figure 9 is a table of model N1 results using the apparatus shown in Figure 1;

Figure 10 is a table of model N2 results using the apparatus shown in Figure 1;

Figure 11 is a table of model N3 results using the apparatus shown in Figure 20   1; and

Figure 12 is a table of classification prediction confidence levels using the apparatus shown in Figure 1.

DESCRIPTION OF THE PREFERRED EMBODIMENT

Referring initially to Figure 1 there is shown an apparatus 10 for classifying 25   media or information flowing through a path of a communications network which in this case is the Internet.

As can be seen, network traffic passing through the apparatus 10 is captured and analysed for providing statistics relating to interactions between two or more terminals (not shown). The captured traffic is first checked against a list of 30   predetermined classifications to determine if it is known or unknown.

6

When the captured traffic is of an unknown classification, various models (to be described more fully below) are applied to the data set in the captured traffic in order to predict the content classification. The models use parameters derived from a knowledge base of previously classified data sets and fitness with these

5   parameters to determine the classification of the content of the newly captured traffic. Thus, the web site sending the captured traffic is now classified and is added to the list of known classifications.

It should be noted that the embodiment of the apparatus 10 as described herein is for an analysis of transmission traffic using the HTTP protocol. The

10   apparatus 10 according to the present invention is not restricted to HTTP, and is easily adaptable to analyse data carried within any networks using any known protocol. Examples of the protocols include FTP, SMTP, NNTP, etc.

Following classification the captured data set is stored in the knowledge base. As the knowledge base expands, more data are used for the model

15   parameters. This refines the apparatus and results in improved predictive performance.

The sites that are deemed to include undesirable information are added to Access control lists (ACLs). The ACLs are used control the flow of content information between terminals. E.g. Undesired content information can be

20   prevented from travelling further through the network by simply not forwarding it, or by replacing it, or by intercepting the request for such content information and modifying its destination.

Classification of traffic from content servers are relatively static. On the other hand, user terminals that interact with these content servers are variable and their

25   classifications are considered transient classifications.

Whereas classifications of content servers form a model of the style of content residing on the server, transient classifications form a model of style of content being viewed by a user terminal, or content consumer. This in effect forms a behaviour profile of such a consumer. This profile can be used to tailor the

30   content information to suit the consumer.

7

As mentioned earlier the apparatus 10 captures a set of observed data relating to a network interaction event, and provides a set of results indicating the classification of a resource or personality residing at each network node involved in the interaction. This is accomplished by applying various statistical models to a

5   profile, and testing this against results obtained from profiles of known classifications. In this example of the invention this process is represented by the following formulas:

x is an unknown profile to be classified;

Profiles $p_1, p_2, p_3...p_n$ are of known classifications;

10   Models $M_1, M_2, M_3...M_n$ are available to operate on these profiles; and

$C_1, C_2, C_3...C_n$ are profile classifications.

The population of a profile of classification C_1, may be defined by the population of $M_1(p)$. $M_1(x)$ may be tested against the true population using any of the standard statistical hypothesis methods.

15   A pre-determined set of media terminals of a classification are modelled by various models $M_1, M_2 .. M_n$. Each model consists of an approach and a set of parameter, e.g linear regression, gradient and point of interception, so that for a single classification $M_1(p_1, p_2 .. p_n)$, $M_2(q_1, q_2 .. q_n) .. M_n(r_1, r_2 .. r_n)$ are used to model the population from the classification. The models may be based on

20   mathematical structures, or arbitrary rules.

The models are continually refined as more network traffic passes through the apparatus 10, thereby increasing the population space from which the classifications are computed.

A terminal may be permanently or transitionally defined in relation to a

25   classification. A transitionally defined terminal may move between classifications based on the fitness of the observed traffic to the models of the various classifications.

Figures 2 to 8 are tables of selected data of traffic for testing the profile of data during a network interaction with a content server to determine if it contains

30   media content of a pornographic nature. Assumption is made that profiles for

content servers contain a variable which is the average size of graphical images served.

A normal distribution or similar non-deterministic probability distribution is then used to test the hypothesis that the profile belongs to a population classified as pornographic. In this example, the population of the classification may be defined by the population of N(a,b) where N is the image size and a and b are the mean and variance respectively, based on a normal distribution. The average and standard deviation derived from the observed samples is tested against the true population using standard statistical hypothesis methods.

In some cases this approach may be broadened to encompass analysis of variance methods with multiple dependant variables, to model the characteristics of a site. Traditional ANOVA or regressive techniques may be applied to model the media content.

A variety of traditional deterministic and non-deterministic models may be applied to determine the hypothesis of profile classification. These may be changed or upgraded continually depending on the level of predictive power found. The functionality of models used is not limited to, but can include simple rules-of-thumb, deterministic and non-deterministic probability models, or arbitrary calculations.

The choice of model is primarily dictated by the predictive power of that model against the population in question.

Figures 2 through 8 show examples of basic data set that can be gathered by observing network traffic of a typical interaction between a client browser and a web server.

Figures 9 to 11 illustrate a simple classification model. This model looks at the size, content and relationships of objects being transmitted by a content server. The outcome of this model is to determine if the media being transmitted has pornographic content.

Classification: pornographic

Standard Model:

$N1(a,b)$

Where N1 is the image size, a and b are the mean and variance respectively, based on a normal distribution.

N2(c,d)

Where N2 is the ratio of text to graphics, c and d are the total size of the text and graphic objects respectively.

N3(e)

Where N3 is the count of word patterns matched from a list of pre-determined words, and e is the text of an object.

Observed Samples are given in the tables shown in Figures 2 to 8.

For model N1 shown in Figure 9, there is applied the normal distribution hypothesis test to the observed samples deriving the results.

The result shows confidence to the 93% and 87% level for sites 6 and 7 respectively, that the sites belong to a population of pornographic sites. The other samples give much lower confidence levels.

For model N2 shown in Figure 10, a simple rule is used to test if the ratio is below a pre-determined threshold. The results show that sites 2, 4, 6 and 7 are within the threshold rating.

For Model N3 shown in Figure 11, a simple rule is used to test if the number of words matching a list of patterns, exceeds a pre-determined threshold. The results show that sites 6 and 7 exceed the threshold.

A weighting formula is then applied to derive a final result as shown in Figure 12.

Therefore, using this example model, the apparatus 10 would predict that sites 6 and 7 are probably serving media with pornographic content, whereas sites 1 through 5 probably are not.

The attached appendix shows an example of the set of rules, constants and formulas which determine a confidence prediction based on logistic regression. The rules are defined using "Submodel" and "Model" components to define individial data points, and aggregated data points. These are then referred to in the "ProbabilityAnalyser" equations which use standard predictive formulas.

Whilst the above has been given by way of illustrative example of the present invention many variations and modifications thereto will be apparent to those skilled in the art without departing from the broad ambit and scope of the invention as herein set forth.

11

CLAIMS

1.   An apparatus for classifying information on communications network, the apparatus comprises means for obtaining one or more transmission characteristics of information on a path of said communications network, and analysing means for predicting a classification of said information based on said one or more transmission characteristics.

2.   An apparatus for classifying content servers which are accessible on a communications network, the apparatus comprises means for obtaining one or more transmission characteristics of information provided by any of said content servers on a path of said communications network, and analysing means for predicting a classification of said information based on said one or more transmission characteristics.

3.   A computer program for classifying information which is accessible on a communications network, the program comprises means for obtaining one or more transmission characteristics of information on a path of said communications network, and analysing means for predicting a classification of said information based on said one or more transmission characteristics.

4.   A computer program for classifying content servers which are accessible on a communications network, the apparatus comprises means for obtaining one or more transmission characteristics of information provided by any of said content servers on a path of said communications network, analysing means for predicting a classification of said information based on said one or more transmission characteristics.

5.   An apparatus for classifying user profiles of users accessing information or content servers on a communications network, the apparatus comprises means for obtaining one or more transmission characteristics of information or information provided by any one of said content servers on a path of said communications network, analysing means for predicting a classification of said information or said one content server based on said one or more transmission characteristics, and means for classifying user profile in accordance with the predicted classification.

12

6.   A computer program for classifying user profiles of users accessing information or content servers on a communications network, the program comprises means for obtaining one or more transmission characteristics of information or information provided by any one of said content servers on a path of said communications network, analysing means for predicting a classification of said information or said one content server based on said one or more transmission characteristics, and means for classifying user profile in accordance with the predicted classification.

7.   The invention according to any one of claims 1 to 6 further comprising means for storing said one or more transmission characteristics.

8.   The invention according to any one of claims 1 to 7 wherein said one or more transmission characteristics include any one or more of network protocol, date and time stamps, size of transmission activities (text and image), content type of transmission activities, pattern seen within the content of the transmission and any other characteristic that can be employed for predicting classifications.

9.   The invention according to any one of claims 1 to 8 wherein said one or more transmission characteristics are obtained from network packets or fragments thereof.

10.   The invention according to any one of claims 1 to 9 wherein the analysing means includes profiling means for providing profiles of interactions based on said one or more transmission characteristics.

11.   The invention according to claim 10 said profiling means is arranged to process said one or more transmission characteristics for providing any one or more of frequency of interaction, duration of interaction, duration of absence of interaction, patterns of transmission, average number of http links within an object of related sites, average number of like sites visited within a time frame, and statistics from said other characteristics, for forming interaction profiles, and the analysing means is adapted to use the profiles for predicting classifications.

12.   The invention according to any one of claims 1 to 11 further comprising a knowledge base of predetermined profiles, and the analysing means is adapted to

predict a classification based on a comparison between the profile of information to be classified and predetermined profiles.

13.   The invention according to claim 12 further comprising means for updating the knowledge base so that the classification prediction can be enhanced following classifications.

14

Appendix

```
      #----------------------------------------------------------
      #Body Text Word Ratio Models
5     SubModel Param AllWordCount WordList AllWords
      SubModel Param AllWordCount Context BODY
      #----------------------------------------------------------
      #Body Text Unique Word Ratio Models
      SubModel Param AllWordCountUnique WordList AllWords
10    SubModel Param AllWordCountUnique Context BODY
      SubModel Param AllWordCountUnique Mode Distinct
      #----------------------------------------------------------
      #Meta Text Word Ratio Models
      SubModel Param AllMetaWordCount WordList AllWords
15    SubModel Param AllMetaWordCount Context META
      #----------------------------------------------------------
      #Alternate Text Word Ratio Models
      SubModel Param AllAlternateWordCount WordList AllWords
      SubModel Param AllAlternateWordCount Context ALTERNATE
20    #----------------------------------------------------------
      #Image models
      SubModel Param LargeGIFPictureCount Dimension 201 x 201 - 999 x 999
      SubModel Param LargeGIFPictureCount ImageType GIF
      SubModel Param ThumbnailGIFPictureCount Dimension 51 x 51 - 200 x 200
25    SubModel Param ThumbnailGIFPictureCount ImageType GIF
      SubModel Param IconGIFPictureCount Dimension 5 x 5 - 50 x 50
      SubModel Param IconGIFPictureCount ImageType GIF
      SubModel Param AllGIFPictureCount ImageType GIF
      Model Exp LargeGIFPictureRatio RATIO(LargeGIFPictureCount,
30    AllGIFPictureCount)
```

Model Exp ThumbnailGIFPictureRatio RATIO(ThumbnailGIFPictureCount,

AllGIFPictureCount)

Model Exp IconGIFPictureRatio RATIO(IconGIFPictureCount,

5   AllGIFPictureCount)

#————

SubModel Param LargeJPEGPictureCount Dimension 201 x 201 - 999 x 999

SubModel Param LargeJPEGPictureCount ImageType JPEG

SubModel Param ThumbnailJPEGPictureCount Dimension 51 x 51 - 200 x 200

10   SubModel Param ThumbnailJPEGPictureCount ImageType JPEG

SubModel Param IconJPEGPictureCount Dimension 5 x 5 - 50 x 50

SubModel Param IconJPEGPictureCount ImageType JPEG

SubModel Param AllJPEGPictureCount ImageType JPEG

Model Exp LargeJPEGPictureRatio RATIO(LargeJPEGPictureCount,

15   AllJPEGPictureCount)

Model Exp ThumbnailJPEGPictureRatio RATIO(ThumbnailJPEGPictureCount,

AllJPEGPictureCount)

Model Exp IconJPEGPictureRatio RATIO(IconJPEGPictureCount,

AllJPEGPictureCount)

20   #————

SubModel Param LowDepthGIFPictureCount Depth 2 - 4

SubModel Param LowDepthGIFPictureCount ImageType GIF

SubModel Param MediumDepthGIFPictureCount Depth 5 - 6

SubModel Param MediumDepthGIFPictureCount ImageType GIF

25   SubModel Param HighDepthGIFPictureCount Depth 7 - 16

SubModel Param HighDepthGIFPictureCount ImageType GIF

Model Exp LowDepthGIFPictureRatio RATIO(LowDepthGIFPictureCount,

AllGIFPictureCount)

Model Exp MediumDepthGIFPictureRatio

30   RATIO(MediumDepthGIFPictureCount, AllGIFPictureCount)

Model Exp HighDepthGIFPictureRatio RATIO(HighDepthGIFPictureCount,

AllGIFPictureCount)

#--------

5   SubModel Param LowDepthJPEGPictureCount Depth 2 - 7

SubModel Param LowDepthJPEGPictureCount ImageType JPEG

SubModel Param MediumDepthJPEGPictureCount Depth 8 - 15

SubModel Param MediumDepthJPEGPictureCount ImageType JPEG

SubModel Param HighDepthJPEGPictureCount Depth 16 - 36

10   SubModel Param HighDepthJPEGPictureCount ImageType JPEG

Model Exp LowDepthJPEGPictureRatio RATIO(LowDepthJPEGPictureCount,

AllJPEGPictureCount)

Model Exp MediumDepthJPEGPictureRatio

RATIO(MediumDepthJPEGPictureCount, AllJPEGPictureCount)

15   Model Exp HighDepthJPEGPictureRatio RATIO(HighDepthJPEGPictureCount,

AllJPEGPictureCount)

#--

#Links Out Models

SubModel Param AllLinkOutCount IncludeLocal FALSE

20   #--------

SubModel Param AVSLinkOutCount Classification ADULTVERIFICATION

SubModel Param AVSLinkOutCount IncludeLocal FALSE

Model Exp AVSLinkOutRatio RATIO(AVSLinkOutCount, AllLinkOutCount)

25   # begin porn.conf

#Body Text Word Count Models

SubModel Param PornExtraHardWordCount WordFile

models/dictionary/porn/porn_words_extrahard.txt

SubModel Param PornHardWordCount WordFile

30   models/dictionary/porn/porn_words_hard.txt

SubModel Param PornMediumWordCount WordFile

models/dictionary/porn/porn_words_medium.txt

SubModel Param PornLiteWordCount WordFile

5   models/dictionary/porn/porn_words_lite.txt

SubModel Param PornExtraLiteWordCount WordFile

models/dictionary/porn/porn_words_extralite.txt

#——

#Unique Body Text Word Count Models

10

SubModel Param PornExtraHardWordCountUnique WordFile

models/dictionary/porn/porn_words_extrahard.txt

SubModel Param PornExtraHardWordCountUnique Mode Distinct

SubModel Param PornHardWordCountUnique WordFile

15   models/dictionary/porn/porn_words_hard.txt

SubModel Param PornHardWordCountUnique Mode Distinct

SubModel Param PornMediumWordCountUnique WordFile

models/dictionary/porn/porn_words_medium.txt

SubModel Param PornMediumWordCountUnique Mode Distinct

20   SubModel Param PornLiteWordCountUnique WordFile

models/dictionary/porn/porn_words_lite.txt

SubModel Param PornLiteWordCountUnique Mode Distinct

SubModel Param PornExtraLiteWordCountUnique WordFile

models/dictionary/porn/porn_words_extralite.txt

25   SubModel Param PornExtraLiteWordCountUnique Mode Distinct

#——

#Body Text Word Ratio Models

Model Exp PornTextWordRatioExtraHard RATIO(PornExtraHardWordCount,

AllWordCount)

30   Model Exp PornTextWordRatioHard RATIO(PornHardWordCount,

AllWordCount)

18

Model Exp PornTextWordRatioMedium RATIO(PornMediumWordCount,

AllWordCount)

Model Exp PornTextWordRatioLite RATIO(PornLiteWordCount, AllWordCount)

5   Model Exp PornTextWordRatioExtraLite RATIO(PornExtraLiteWordCount,

AllWordCount)

        #--

        #Body Text Unique Word Ratio Models

        Model Exp PornTextWordRatioExtraHardUnique

10   RATIO(PornExtraHardWordCountUnique, AllWordCountUnique)

        Model Exp PornTextWordRatioHardUnique

        RATIO(PornHardWordCountUnique, AllWordCountUnique)

        Model Exp PornTextWordRatioMediumUnique

15   RATIO(PornMediumWordCountUnique, AllWordCountUnique)

        Model Exp PornTextWordRatioLiteUnique RATIO(PornLiteWordCountUnique,

        AllWordCountUnique)

        Model Exp PornTextWordRatioExtraLiteUnique

        RATIO(PornExtraLiteWordCountUnique, AllWordCountUnique)

20   #--

        #Domain Word Count Models

        SubModel Param PornExtraHardDomainWordCount Context DOMAIN-NAME

        SubModel Param PornExtraHardDomainWordCount WordFile

        models/dictionary/porn/porn_words_extrahard.txt

25   SubModel Param PornHardDomainWordCount Context DOMAIN-NAME

        SubModel Param PornHardDomainWordCount WordFile

        models/dictionary/porn/porn_words_hard.txt

        SubModel Param PornMediumDomainWordCount Context DOMAIN-NAME

        SubModel Param PornMediumDomainWordCount WordFile

30   models/dictionary/porn/porn_words_medium.txt

        SubModel Param PornLiteDomainWordCount Context DOMAIN-NAME

19

SubModel Param PornLiteDomainWordCount WordFile

models/dictionary/porn/porn_words_lite.txt

SubModel Param PornExtraLiteDomainWordCount Context DOMAIN-NAME

5   SubModel Param PornExtraLiteDomainWordCount WordFile

models/dictionary/porn/porn_words_extralite.txt

#---

#Meta Text Word Count Models

SubModel Param PornExtraHardMetaWordCount Context META

10   SubModel Param PornExtraHardMetaWordCount WordFile

models/dictionary/porn/porn_words_extrahard.txt

SubModel Param PornHardMetaWordCount Context META

SubModel Param PornHardMetaWordCount WordFile

models/dictionary/porn/porn_words_hard.txt

15

SubModel Param PornMediumMetaWordCount Context META

SubModel Param PornMediumMetaWordCount WordFile

models/dictionary/porn/porn_words_medium.txt

SubModel Param PornLiteMetaWordCount Context META

20   SubModel Param PornLiteMetaWordCount WordFile

models/dictionary/porn/porn_words_lite.txt

SubModel Param PornExtraLiteMetaWordCount Context META

SubModel Param PornExtraLiteMetaWordCount WordFile

models/dictionary/porn/porn_words_extralite.txt

25   #---

#Meta Text Word Ratio Models

Model Exp PornMetaWordRatioExtraHard

RATIO(PornExtraHardMetaWordCount, AllMetaWordCount)

Model Exp PornMetaWordRatioHard RATIO(PornHardMetaWordCount,

30   AllMetaWordCount)

Model Exp PornMetaWordRatioMedium RATIO(PornMediumMetaWordCount,

AllMetaWordCount)

Model Exp PornMetaWordRatioLite RATIO(PornLiteMetaWordCount,

5   AllMetaWordCount)

Model Exp PornMetaWordRatioExtraLite RATIO(PornExtraLiteMetaWordCount,

AllMetaWordCount)

#————————————— —————————————————

#Alternate Text Word Count Models

10   SubModel Param PornExtraHardAlternateWordCount Context ALTERNATE

SubModel Param PornExtraHardAlternateWordCount WordFile

models/dictionary/porn/porn_words_extrahard.txt

SubModel Param PornHardAlternateWordCount Context ALTERNATE

SubModel Param PornHardAlternateWordCount WordFile

15   models/dictionary/porn/porn_words_hard.txt

SubModel Param PornMediumAlternateWordCount Context ALTERNATE

SubModel Param PornMediumAlternateWordCount WordFile

models/dictionary/porn/porn_words_medium.txt

SubModel Param PornLiteAlternateWordCount Context ALTERNATE

20   SubModel Param PornLiteAlternateWordCount WordFile

models/dictionary/porn/porn_words_lite.txt

SubModel Param PornExtraLiteAlternateWordCount Context ALTERNATE

SubModel Param PornExtraLiteAlternateWordCount WordFile

models/dictionary/porn/porn_words_extralite.txt

25   #——————————————————————————————————————

#Alternate Text Word Ratio Models

Model Exp PornAlternateWordRatioExtraHard

RATIO(PornExtraHardAlternateWordCount, AllAlternateWordCount)

Model Exp PornAlternateWordRatioHard RATIO(PornHardAlternateWordCount,

30   AllAlternateWordCount)

Model Exp PornAlternateWordRatioMedium

RATIO(PornMediumAlternateWordCount, AllAlternateWordCount)

Model Exp PornAlternateWordRatioLite RATIO(PornLiteAlternateWordCount,

5   AllAlternateWordCount)

Model Exp PornAlternateWordRatioExtraLite

RATIO(PornExtraLiteAlternateWordCount, AllAlternateWordCount)

#--

#Links Out Models

10   SubModel Param PornLinkOutCount Classification PORN

SubModel Param PornLinkOutCount IncludeLocal FALSE

Model Exp PornLinkOutRatio RATIO(PornLinkOutCount, AllLinkOutCount)

#--

#Logistic Models

15   Model Exp PornLRConstant -3.9869

#--------

Model Exp PornLRCoefficientPornTextWordRatioExtraHard 39.7450

Model Exp PornLRCoefficientPornTextWordRatioHard 355.0550

20   Model Exp PornLRCoefficientPornTextWordRatioMedium -136.436

Model Exp PornLRCoefficientPornTextWordRatioLite -63.2565

Model Exp PornLRCoefficientPornTextWordRatioExtraLite 33.9054

#--------

Model Exp PornLRCoefficientPornTextWordRatioExtraHardUnique 111.4752

25   Model Exp PornLRCoefficientPornTextWordRatioHardUnique -72.7005

Model Exp PornLRCoefficientPornTextWordRatioMediumUnique 264.1902

Model Exp PornLRCoefficientPornTextWordRatioLiteUnique 125.0743

Model Exp PornLRCoefficientPornTextWordRatioExtraLiteUnique -16.6895

#--------

30   Model Exp PornLRCoefficientPornExtraHardDomainWordCount 0.2598

Model Exp PornLRCoefficientPornHardDomainWordCount 2.1344

22

Model Exp PornLRCoefficientPornMediumDomainWordCount 0

Model Exp PornLRCoefficientPornLiteDomainWordCount 0.0610

Model Exp PornLRCoefficientPornExtraLiteDomainWordCount 0

5　　#———

Model Exp PornLRCoefficientPornMetaWordRatioExtraHard 0

Model Exp PornLRCoefficientPornMetaWordRatioHard 0

Model Exp PornLRCoefficientPornMetaWordRatioMedium 0

Model Exp PornLRCoefficientPornMetaWordRatioLite 0

10　　Model Exp PornLRCoefficientPornMetaWordRatioExtraLite 0

#———

Model Exp PornLRCoefficientPornAlternateWordRatioExtraHard 16.1972

Model Exp PornLRCoefficientPornAlternateWordRatioHard 0

Model Exp PornLRCoefficientPornAlternateWordRatioMedium 26.4186

15　　Model Exp PornLRCoefficientPornAlternateWordRatioLite 0

Model Exp PornLRCoefficientPornAlternateWordRatioExtraLite 14.1615

#———

Model Exp PornLRCoefficientAllGIFPictureCount 0

Model Exp PornLRCoefficientLargeGIFPictureCount 0

20

Model Exp PornLRCoefficientIconGIFPictureCount 0

Model Exp PornLRCoefficientThumbnailGIFPictureCount 0

Model Exp PornLRCoefficientLargeGIFPictureRatio 0

Model Exp PornLRCoefficientIconGIFPictureRatio 0

25　　Model Exp PornLRCoefficientThumbnailGIFPictureRatio 0

Model Exp PornLRCoefficientHighDepthGIFPictureCount 0

Model Exp PornLRCoefficientMediumDepthGIFPictureCount 0

Model Exp PornLRCoefficientLowDepthGIFPictureCount 0

Model Exp PornLRCoefficientHighDepthGIFPictureRatio 0

30　　Model Exp PornLRCoefficientMediumDepthGIFPictureRatio 0

Model Exp PornLRCoefficientLowDepthGIFPictureRatio 0

Model Exp PornLRCoefficientAllJPEGPictureCount 0

Model Exp PornLRCoefficientLargeJPEGPictureCount 0

5   Model Exp PornLRCoefficientIconJPEGPictureCount 0

Model Exp PornLRCoefficientThumbnailJPEGPictureCount 0

Model Exp PornLRCoefficientLargeJPEGPictureRatio 0

Model Exp PornLRCoefficientIconJPEGPictureRatio 0

Model Exp PornLRCoefficientThumbnailJPEGPictureRatio 0

10   Model Exp PornLRCoefficientHighDepthJPEGPictureCount 0

Model Exp PornLRCoefficientMediumDepthJPEGPictureCount 0

Model Exp PornLRCoefficientLowDepthJPEGPictureCount 0

Model Exp PornLRCoefficientHighDepthJPEGPictureRatio 0

Model Exp PornLRCoefficientMediumDepthJPEGPictureRatio 0

15   Model Exp PornLRCoefficientLowDepthJPEGPictureRatio 0

#————

Model Exp PornLRCoefficientPornLinkOutRatio 4.6958

Model Exp PornLRCoefficientAVSLinkOutCount 0.3327

Model Exp PornLRCoefficientAVSLinkOutRatio 3.6786

20   #————

Model Exp PornLRLogOdds SUM(PornLRConstant, \

       PRODUCT(PornLRCoefficientPornTextWordRatioExtraHard,

PornTextWordRatioExtraHard), \

       PRODUCT(PornLRCoefficientPornTextWordRatioHard,

25   PornTextWordRatioHard), \

       PRODUCT(PornLRCoefficientPornTextWordRatioMedium,

PornTextWordRatioMedium), \

       PRODUCT(PornLRCoefficientPornTextWordRatioLite,

PornTextWordRatioLite), \

30   PRODUCT(PornLRCoefficientPornTextWordRatioExtraLite,

PornTextWordRatioExtraLite), \

24

PRODUCT(PornLRCoefficientPornTextWordRatioExtraHardUnique,

PornTextWordRatioExtraHardUnique), \

PRODUCT(PornLRCoefficientPornTextWordRatioHardUnique,

5   PornTextWordRatioHardUnique), \

PRODUCT(PornLRCoefficientPornTextWordRatioMediumUnique,

PornTextWordRatioMediumUnique), \

PRODUCT(PornLRCoefficientPornTextWordRatioLiteUnique,

PornTextWordRatioLiteUnique), \

10   PRODUCT(PornLRCoefficientPornTextWordRatioExtraLiteUnique,

PornTextWordRatioExtraLiteUnique), \

PRODUCT(PornLRCoefficientPornExtraHardDomainWordCount,

PornExtraHardDomainWordCount), \

PRODUCT(PornLRCoefficientPornHardDomainWordCount,

15   PornHardDomainWordCount), \

PRODUCT(PornLRCoefficientPornMediumDomainWordCount,

PornMediumDomainWordCount), \

PRODUCT(PornLRCoefficientPornLiteDomainWordCount,

PornLiteDomainWordCount), \

20   PRODUCT(PornLRCoefficientPornExtraLiteDomainWordCount,

PornExtraLiteDomainWordCount), \

PRODUCT(PornLRCoefficientPornMetaWordRatioExtraHard,

PornMetaWordRatioExtraHard), \

PRODUCT(PornLRCoefficientPornMetaWordRatioHard,

25   PornMetaWordRatioHard), \

PRODUCT(PornLRCoefficientPornMetaWordRatioMedium,

PornMetaWordRatioMedium), \

PRODUCT(PornLRCoefficientPornMetaWordRatioLite,

PornMetaWordRatioLite), \

30   PRODUCT(PornLRCoefficientPornMetaWordRatioExtraLite,

PornMetaWordRatioExtraLite), \

```
        PRODUCT(PornLRCoefficientPornAlternateWordRatioExtraHard,
PornAlternateWordRatioExtraHard), \
        PRODUCT(PornLRCoefficientPornAlternateWordRatioHard,
PornAlternateWordRatioHard), \
        PRODUCT(PornLRCoefficientPornAlternateWordRatioMedium,
PornAlternateWordRatioMedium), \
        PRODUCT(PornLRCoefficientPornAlternateWordRatioLite,
PornAlternateWordRatioLite), \
        PRODUCT(PornLRCoefficientPornAlternateWordRatioExtraLite,
PornAlternateWordRatioExtraLite), \
        PRODUCT(PornLRCoefficientAllGIFPictureCount, AllGIFPictureCount), \
        PRODUCT(PornLRCoefficientLargeGIFPictureCount,
LargeGIFPictureCount), \
        PRODUCT(PornLRCoefficientIconGIFPictureCount,
IconGIFPictureCount), \
        PRODUCT(PornLRCoefficientThumbnailGIFPictureCount,
ThumbnailGIFPictureCount), \
        PRODUCT(PornLRCoefficientLargeGIFPictureRatio,
LargeGIFPictureRatio), \
        PRODUCT(PornLRCoefficientIconGIFPictureRatio, IconGIFPictureRatio), \
        PRODUCT(PornLRCoefficientThumbnailGIFPictureRatio,
ThumbnailGIFPictureRatio), \
        PRODUCT(PornLRCoefficientHighDepthGIFPictureCount,
HighDepthGIFPictureCount), \
        PRODUCT(PornLRCoefficientMediumDepthGIFPictureCount,
MediumDepthGIFPictureCount), \
        PRODUCT(PornLRCoefficientLowDepthGIFPictureCount,
LowDepthGIFPictureCount), \
        PRODUCT(PornLRCoefficientHighDepthGIFPictureRatio,
HighDepthGIFPictureRatio), \
```

26

PRODUCT(PornLRCoefficientMediumDepthGIFPictureRatio,

MediumDepthGIFPictureRatio), \

PRODUCT(PornLRCoefficientLowDepthGIFPictureRatio,

5   LowDepthGIFPictureRatio), \

PRODUCT(PornLRCoefficientAllJPEGPictureCount, AllJPEGPictureCount),

\

PRODUCT(PornLRCoefficientLargeJPEGPictureCount,

LargeJPEGPictureCount), \

10   PRODUCT(PornLRCoefficientIconJPEGPictureCount,

IconJPEGPictureCount), \

PRODUCT(PornLRCoefficientThumbnailJPEGPictureCount,

ThumbnailJPEGPictureCount), \

PRODUCT(PornLRCoefficientLargeJPEGPictureRatio,

15   LargeJPEGPictureRatio), \

PRODUCT(PornLRCoefficientIconJPEGPictureRatio,

IconJPEGPictureRatio), \

PRODUCT(PornLRCoefficientThumbnailJPEGPictureRatio,

ThumbnailJPEGPictureRatio), \

20   PRODUCT(PornLRCoefficientHighDepthJPEGPictureCount,

HighDepthJPEGPictureCount), \

PRODUCT(PornLRCoefficientMediumDepthJPEGPictureCount,

MediumDepthJPEGPictureCount), \

PRODUCT(PornLRCoefficientLowDepthJPEGPictureCount,

25   LowDepthJPEGPictureCount), \

PRODUCT(PornLRCoefficientHighDepthJPEGPictureRatio,

HighDepthJPEGPictureRatio), \

PRODUCT(PornLRCoefficientMediumDepthJPEGPictureRatio,

MediumDepthJPEGPictureRatio), \

30   PRODUCT(PornLRCoefficientLowDepthJPEGPictureRatio,

LowDepthJPEGPictureRatio), \

```
            PRODUCT(PornLRCoefficientPornLinkOutRatio, PornLinkOutRatio), \
            PRODUCT(PornLRCoefficientAVSLinkOutCount, AVSLinkOutCount), \
            PRODUCT(PornLRCoefficientAVSLinkOutRatio, AVSLinkOutRatio))
5   #-------------------------------------------------------------
    #Probability Analysers
    ProbabilityAnalyser Param PornAltMetaWordCountProbability Classification
    PORN
    ProbabilityAnalyser Exp PornAltMetaWordCountProbability \
10          SUM(PornExtraHardMetaWordCount, PornHardMetaWordCount, \
                PRODUCT(0.5,PornMediumMetaWordCount), \
                PornExtraHardAlternateWordCount,
    PornHardAlternateWordCount, \
                PRODUCT(0.5,PornMediumAlternateWordCount))
15  ProbabilityAnalyser Param PornMetaWordRatioProbability Classification PORN
    ProbabilityAnalyser Exp PornMetaWordRatioProbability \
            PRODUCT(100, SUM(PornMetaWordRatioExtraHard, \
                PornMetaWordRatioHard, PornMetaWordRatioMedium))
    ProbabilityAnalyser Param PornLRProbability Classification PORN
20  ProbabilityAnalyser Exp PornLRProbability PRODUCT(100,
    RATIO(1,SUM(1,EXP(MINUS(PornLRLogOdds)))))
    #-------------------------------------------------------------
```

1 / 9
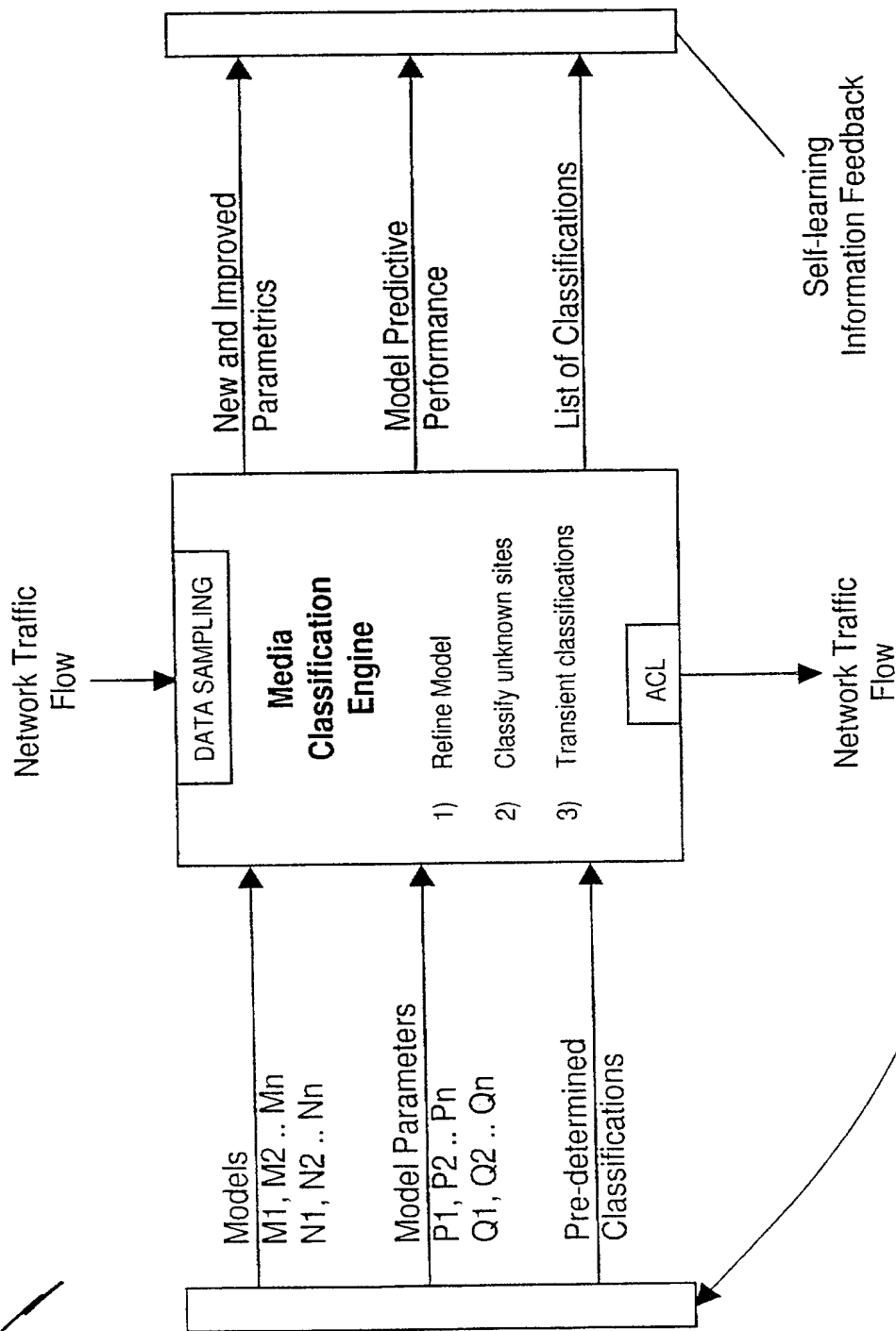
FIG. 1

2 / 9

Site 1 Objects - Content Type "Search Engine"							
	Source		Destination				
Object	IP Address	Port	IP Address	Port	Type	Size	Timestamp
1	202.139.16.45	63450	204.71.200.72	80	GET	36	14:53:17
2	204.71.200.72	80	202.139.16.45	63450	text/html	9424	14:53:17
3	202.139.16.45	63450	204.71.200.72	80	GET	79	14:53:19
4	202.139.16.45	63450	204.71.200.72	80	GET	46	14:53:19
5	202.139.16.45	63450	204.71.200.72	80	GET	50	14:53:19
6	204.71.200.72	80	202.139.16.45	63450	image/gif	2637	14:53:19
7	204.71.200.72	80	202.139.16.45	63450	image/gif	4672	14:53:20
8	204.71.200.72	80	202.139.16.45	63450	image/gif	357	14:53:20
9	202.139.16.45	63450	204.71.200.72	80	GET	56	14:53:27
10	204.71.200.72	80	202.139.16.45	63450	text/html	11193	14:53:28
11	202.139.16.45	63450	204.71.200.72	80	GET	59	14:53:29
12	204.71.200.72	80	202.139.16.45	63450	image/gif	11522	14:53:29
13	202.139.16.45	63450	204.71.200.72	80	GET	67	14:53:30
14	202.139.16.45	63450	204.71.200.72	80	GET	67	14:53:30
15	202.139.16.45	63450	204.71.200.72	80	GET	59	14:53:30
16	204.71.200.72	80	202.139.16.45	63450	image/gif	1398	14:53:30
17	202.139.16.45	63450	204.71.200.72	80	GET	76	14:53:30
18	204.71.200.72	80	202.139.16.45	63450	image/gif	1728	14:53:30
19	202.139.16.45	63450	204.71.200.72	80	GET	69	14:53:30
20	204.71.200.72	80	202.139.16.45	63450	image/gif	962	14:53:30
21	204.71.200.72	80	202.139.16.45	63450	image/gif	946	14:53:31
22	204.71.200.72	80	202.139.16.45	63450	image/gif	1716	14:53:31

FIG. 2

Object	Source IP Address	Source Port	Destination IP Address	Destination Port	Type	Size	Timestamp
	Source		Destination		Type	Size	Timestamp
	IP Address	Port	IP Address	Port			
1	202.139.16.45	63450	165.69.1.187	80	GET	38	14:54:04
2	165.69.1.187	80	202.139.16.45	63450	text/html	2312	14:54:04
3	202.139.16.45	63450	165.69.1.187	80	GET	52	14:54:05
4	202.139.16.45	63450	165.69.1.187	80	GET	56	14:54:05
5	202.139.16.45	63450	165.69.1.187	80	GET	46	14:54:05
6	202.139.16.45	63450	165.69.1.187	80	GET	48	14:54:05
7	202.139.16.45	63450	165.69.1.187	80	GET	47	14:54:05
8	202.139.16.45	63450	165.69.1.187	80	GET	47	14:54:05
9	165.69.1.187	80	202.139.16.45	63450	text/html	333	14:54:05
10	165.69.1.187	80	202.139.16.45	63450	text/html	56	14:54:06
11	165.69.1.187	80	202.139.16.45	63450	text/html	2445	14:54:06
12	165.69.1.187	80	202.139.16.45	63450	text/html	202	14:54:06
13	165.69.1.187	80	202.139.16.45	63450	text/html	202	14:54:06
14	165.69.1.187	80	202.139.16.45	63450	text/html	56	14:54:06
15	202.139.16.45	63450	165.69.1.187	80	GET	71	14:54:06
16	165.69.1.187	80	202.139.16.45	63450	image/gif	1229	14:54:06
17	202.139.16.45	63450	165.69.1.187	80	GET	64	14:54:06
18	165.69.1.187	80	202.139.16.45	63450	image/gif	43	14:54:06
19	202.139.16.45	63450	165.69.1.187	80	GET	64	14:54:07
20	202.139.16.45	63450	165.69.1.187	80	GET	64	14:54:07
21	165.69.1.187	80	202.139.16.45	63450	image/gif	43	14:54:07
22	165.69.1.187	80	202.139.16.45	63450	image/gif	43	14:54:07
23	202.139.16.45	63450	165.69.1.187	80	GET	67	14:54:07
24	165.69.1.187	80	202.139.16.45	63450	image/gif	2442	14:54:07
25	202.139.16.45	63450	165.69.1.187	80	GET	73	14:54:07
26	165.69.1.187	80	202.139.16.45	63450	image/gif	1364	14:54:07
27	202.139.16.45	63450	165.69.1.187	80	GET	71	14:54:07
28	165.69.1.187	80	202.139.16.45	63450	image/gif	8942	14:54:07
29	202.139.16.45	63450	165.69.1.187	80	GET	71	14:54:07
30	202.139.16.45	63450	165.69.1.187	80	GET	65	14:54:08
31	165.69.1.187	80	202.139.16.45	63450	unknown	10	14:54:08
32	165.69.1.187	80	202.139.16.45	63450	image/gif	15550	14:54:08
33	202.139.16.45	63450	165.69.1.187	80	GET	77	14:54:08
34	165.69.1.187	80	202.139.16.45	63450	image/gif	4732	14:54:08
35	202.139.16.45	63450	165.69.1.187	80	GET	70	14:54:08
36	202.139.16.45	63450	165.69.1.187	80	GET	70	14:54:09
37	202.139.16.45	63450	165.69.1.187	80	GET	68	14:54:09
38	202.139.16.45	63450	165.69.1.187	80	GET	68	14:54:09
39	165.69.1.187	80	202.139.16.45	63450	image/gif	436	14:54:09
40	165.69.1.187	80	202.139.16.45	63450	image/gif	405	14:54:09
41	165.69.1.187	80	202.139.16.45	63450	image/gif	436	14:54:09
42	165.69.1.187	80	202.139.16.45	63450	image/gif	405	14:54:09

FIG. 3

3/9

Object	Source IP Address	Port	Destination IP Address	Port	Type	Size	Timestamp
			Site 2 Objects - Content Type "News"				
1	202.139.16.45	63450	165.69.1.187	80	GET	38	14:54:04
2	165.69.1.187	80	202.139.16.45	63450	text/html	2312	14:54:04
3	202.139.16.45	63450	165.69.1.187	80	GET	52	14:54:05
4	202.139.16.45	63450	165.69.1.187	80	GET	56	14:54:05
5	202.139.16.45	63450	165.69.1.187	80	GET	46	14:54:05
6	202.139.16.45	63450	165.69.1.187	80	GET	48	14:54:05
7	202.139.16.45	63450	165.69.1.187	80	GET	47	14:54:05
8	202.139.16.45	63450	165.69.1.187	80	GET	47	14:54:05
9	165.69.1.187	80	202.139.16.45	63450	text/html	333	14:54:05
10	165.69.1.187	80	202.139.16.45	63450	text/html	56	14:54:06
11	165.69.1.187	80	202.139.16.45	63450	text/html	2445	14:54:06
12	165.69.1.187	80	202.139.16.45	63450	text/html	202	14:54:06
13	165.69.1.187	80	202.139.16.45	63450	text/html	202	14:54:06
14	165.69.1.187	80	202.139.16.45	63450	text/html	56	14:54:06
15	202.139.16.45	63450	165.69.1.187	80	GET	71	14:54:06
16	165.69.1.187	80	202.139.16.45	63450	image/gif	1229	14:54:06
17	202.139.16.45	63450	165.69.1.187	80	GET	64	14:54:06
18	165.69.1.187	80	202.139.16.45	63450	image/gif	43	14:54:06
19	202.139.16.45	63450	165.69.1.187	80	GET	64	14:54:06
20	202.139.16.45	63450	165.69.1.187	80	GET	64	14:54:07
21	165.69.1.187	80	202.139.16.45	63450	image/gif	43	14:54:07
22	165.69.1.187	80	202.139.16.45	63450	image/gif	43	14:54:07
23	202.139.16.45	63450	165.69.1.187	80	GET	67	14:54:07
24	165.69.1.187	80	202.139.16.45	63450	image/gif	2442	14:54:07
25	202.139.16.45	63450	165.69.1.187	80	GET	73	14:54:07
26	165.69.1.187	80	202.139.16.45	63450	image/gif	1364	14:54:07
27	202.139.16.45	63450	165.69.1.187	80	GET	71	14:54:07
28	165.69.1.187	80	202.139.16.45	63450	image/gif	8942	14:54:07
29	202.139.16.45	63450	165.69.1.187	80	GET	71	14:54:07
30	202.139.16.45	63450	165.69.1.187	80	GET	65	14:54:08
31	165.69.1.187	80	202.139.16.45	63450	unknown	10	14:54:08
32	165.69.1.187	80	202.139.16.45	63450	image/gif	15550	14:54:08
33	202.139.16.45	63450	165.69.1.187	80	GET	77	14:54:08
34	165.69.1.187	80	202.139.16.45	63450	image/gif	4732	14:54:08
35	202.139.16.45	63450	165.69.1.187	80	GET	70	14:54:08
36	202.139.16.45	63450	165.69.1.187	80	GET	70	14:54:09
37	202.139.16.45	63450	165.69.1.187	80	GET	68	14:54:09
38	202.139.16.45	63450	165.69.1.187	80	GET	68	14:54:09
39	165.69.1.187	80	202.139.16.45	63450	image/gif	436	14:54:09
40	165.69.1.187	80	202.139.16.45	63450	image/gif	405	14:54:09
41	165.69.1.187	80	202.139.16.45	63450	image/gif	436	14:54:09
42	165.69.1.187	80	202.139.16.45	63450	image/gif	405	14:54:09

FIG. 3

4/9

Object	Source IP Address	Port	Destination IP Address	Port	Type	Size	Timestamp
			Site 3 Objects - Content Type "Entertainment"				
1	202.139.16.45	63450	204.202.129.23	80	GET	40	14:55:10
2	204.202.129.23	80	202.139.16.45	63450	text/html	56	14:55:11
3	202.139.16.45	63450	204.202.129.23	80	GET	39	14:55:11
4	204.202.129.23	80	202.139.16.45	63450	text/html	25765	14:55:12
5	202.139.16.45	63450	204.202.129.23	80	GET	58	14:55:13
6	202.139.16.45	63450	204.202.129.23	80	GET	66	14:55:13
7	202.139.16.45	63450	204.202.129.23	80	GET	62	14:55:13
8	204.202.129.23	80	202.139.16.45	63450	image/jpeg	5277	14:55:14
9	204.202.129.23	80	202.139.16.45	63450	image/gif	43	14:55:14
10	204.202.129.23	80	202.139.16.45	63450	image/gif	1266	14:55:14
11	202.139.16.45	63450	204.202.129.23	80	GET	63	14:55:20
12	202.139.16.45	63450	204.202.129.23	80	GET	58	14:55:21
13	202.139.16.45	63450	204.202.129.23	80	GET	72	14:55:21
14	204.202.129.23	80	202.139.16.45	63450	image/gif	1733	14:55:22
15	204.202.129.23	80	202.139.16.45	63450	image/gif	5314	14:55:22
16	204.202.129.23	80	202.139.16.45	63450	image/gif	414	14:55:22
17	202.139.16.45	63450	204.202.129.23	80	GET	68	14:55:22
18	202.139.16.45	63450	204.202.129.23	80	GET	62	14:55:22
19	204.202.129.23	80	202.139.16.45	63450	image/gif	406	14:55:22
20	204.202.129.23	80	202.139.16.45	63450	image/gif	746	14:55:22
21	202.139.16.45	63450	204.202.129.23	80	GET	65	14:55:23
22	202.139.16.45	63450	204.202.129.23	80	GET	58	14:55:23
23	202.139.16.45	63450	204.202.129.23	80	GET	63	14:55:23
24	202.139.16.45	63450	204.202.129.23	80	GET	62	14:55:23
25	204.202.129.23	80	202.139.16.45	63450	image/gif	1665	14:55:23
26	204.202.129.23	80	202.139.16.45	63450	image/gif	35	14:55:24
27	204.202.129.23	80	202.139.16.45	63450	image/gif	906	14:55:24
28	204.202.129.23	80	202.139.16.45	63450	image/gif	447	14:55:24
29	202.139.16.45	63450	204.202.129.23	80	GET	67	14:55:24
30	202.139.16.45	63450	204.202.129.23	80	GET	58	14:55:24
31	202.139.16.45	63450	204.202.129.23	80	GET	62	14:55:24
32	204.202.129.23	80	202.139.16.45	63450	image/jpeg	7861	14:55:24
33	204.202.129.23	80	202.139.16.45	63450	image/gif	391	14:55:25
34	204.202.129.23	80	202.139.16.45	63450	image/gif	641	14:55:25
35	202.139.16.45	63450	204.202.129.23	80	GET	57	14:55:25
36	204.202.129.23	80	202.139.16.45	63450	image/gif	377	14:55:25
37	202.139.16.45	63450	204.202.129.23	80	GET	60	14:55:26
38	202.139.16.45	63450	204.202.129.23	80	GET	78	14:55:26
39	202.139.16.45	63450	204.202.129.23	80	GET	74	14:55:26
40	202.139.16.45	63450	204.202.129.23	80	GET	67	14:55:26
41	204.202.129.23	80	202.139.16.45	63450	image/gif	403	14:55:26
42	204.202.129.23	80	202.139.16.45	63450	image/gif	1796	14:55:26
43	204.202.129.23	80	202.139.16.45	63450	image/gif	6845	14:55:27
44	202.139.16.45	63450	204.202.129.23	80	GET	56	14:55:27
45	204.202.129.23	80	202.139.16.45	63450	image/jpeg	17796	14:55:27
46	202.139.16.45	63450	204.202.129.23	80	GET	56	14:55:27
47	204.202.129.23	80	202.139.16.45	63450	image/gif	49	14:55:27
48	204.202.129.23	80	202.139.16.45	63450	image/gif	44	14:55:27

FIG. 4

5 / 9

Site 4 Objects - Content Type "Computer Hardware"						
	Source		Destination			
Object	IP Address	Port	IP Address	Port	Type	Size	Timestamp
1	202.139.16.45	63450	209.143.240.6	80	GET	34	14:55:49
2	209.143.240.6	80	202.139.16.45	63450	text/html	1645	14:55:50
3	202.139.16.45	63450	209.143.240.6	80	GET	47	14:55:51
4	209.143.240.6	80	202.139.16.45	63450	text/html	56	14:55:51
5	202.139.16.45	63450	209.143.240.6	80	GET	64	14:55:53
6	202.139.16.45	63450	209.143.240.6	80	GET	63	14:55:53
7	202.139.16.45	63450	209.143.240.6	80	GET	63	14:55:53
8	209.143.240.6	80	202.139.16.45	63450	image/gif	290	14:55:53
9	202.139.16.45	63450	209.143.240.6	80	GET	64	14:55:54
10	209.143.240.6	80	202.139.16.45	63450	image/gif	403	14:55:54
11	209.143.240.6	80	202.139.16.45	63450	image/gif	381	14:55:54
12	202.139.16.45	63450	209.143.240.6	80	GET	63	14:55:54
13	202.139.16.45	63450	209.143.240.6	80	GET	65	14:55:54
14	209.143.240.6	80	202.139.16.45	63450	image/gif	348	14:55:54
15	202.139.16.45	63450	209.143.240.6	80	GET	65	14:55:54
16	202.139.16.45	63450	209.143.240.6	80	GET	65	14:55:54
17	209.143.240.6	80	202.139.16.45	63450	image/gif	354	14:55:54
18	209.143.240.6	80	202.139.16.45	63450	image/gif	600	14:55:54
19	202.139.16.45	63450	209.143.240.6	80	GET	65	14:55:55
20	202.139.16.45	63450	209.143.240.6	80	GET	54	14:55:55
21	209.143.240.6	80	202.139.16.45	63450	image/gif	490	14:55:55
22	202.139.16.45	63450	209.143.240.6	80	GET	53	14:55:55
23	209.143.240.6	80	202.139.16.45	63450	image/gif	571	14:55:55
24	209.143.240.6	80	202.139.16.45	63450	image/gif	322	14:55:55
25	209.143.240.6	80	202.139.16.45	63450	image/gif	571	14:55:55
26	202.139.16.45	63450	209.143.240.6	80	GET	55	14:55:55
27	209.143.240.6	80	202.139.16.45	63450	image/gif	363	14:55:55
28	202.139.16.45	63450	209.143.240.6	80	GET	63	14:55:55
29	202.139.16.45	63450	209.143.240.6	80	GET	63	14:55:55
30	202.139.16.45	63450	209.143.240.6	80	GET	62	14:55:55
31	209.143.240.6	80	202.139.16.45	63450	image/gif	241	14:55:56
32	202.139.16.45	63450	209.143.240.6	80	GET	59	14:55:56
33	209.143.240.6	80	202.139.16.45	63450	image/gif	488	14:55:56
34	209.143.240.6	80	202.139.16.45	63450	image/gif	463	14:55:56
35	202.139.16.45	63450	209.143.240.6	80	GET	59	14:55:56
36	209.143.240.6	80	202.139.16.45	63450	image/gif	714	14:55:56
37	202.139.16.45	63450	209.143.240.6	80	GET	53	14:55:56
38	209.143.240.6	80	202.139.16.45	63450	image/gif	35	14:55:56
39	202.139.16.45	63450	209.143.240.6	80	GET	53	14:55:56
40	209.143.240.6	80	202.139.16.45	63450	image/gif	1188	14:55:56
41	202.139.16.45	63450	209.143.240.6	80	GET	54	14:55:56
42	209.143.240.6	80	202.139.16.45	63450	image/gif	327	14:55:57

FIG. 5

6 / 9

Site 5 Objects - Content Type "Computer Software"							
	Source		Destination				
Object	IP Address	Port	IP Address	Port	Type	Size	Timestamp
1	202.139.16.45	63450	207.46.130.149	80	GET	40	14:57:46
2	207.46.130.149	80	202.139.16.45	63450	text/html	19019	14:57:47
3	202.139.16.45	63450	207.46.130.149	80	GET	66	14:57:48
4	207.46.130.149	80	202.139.16.45	63450	application	18768	14:57:49
5	202.139.16.45	63450	207.46.130.149	80	GET	71	14:57:50
6	207.46.130.149	80	202.139.16.45	63450	application	2941	14:57:50
7	202.139.16.45	63450	207.46.130.149	80	GET	70	14:57:50
8	207.46.130.149	80	202.139.16.45	63450	application	2812	14:57:51
9	202.139.16.45	63450	207.46.130.149	80	GET	75	14:57:51
10	202.139.16.45	63450	207.46.130.149	80	GET	72	14:57:51
11	202.139.16.45	63450	207.46.130.149	80	GET	80	14:57:51
12	202.139.16.45	63450	207.46.130.149	80	GET	73	14:57:51
13	207.46.130.149	80	202.139.16.45	63450	image/gif	7976	14:57:51
14	202.139.16.45	63450	207.46.130.149	80	GET	80	14:57:52
15	207.46.130.149	80	202.139.16.45	63450	image/gif	139	14:57:52
16	207.46.130.149	80	202.139.16.45	63450	image/gif	44	14:57:52
17	207.46.130.149	80	202.139.16.45	63450	image/gif	666	14:57:52
18	202.139.16.45	63450	207.46.130.149	80	GET	80	14:57:52
19	207.46.130.149	80	202.139.16.45	63450	image/gif	44	14:57:52
20	207.46.130.149	80	202.139.16.45	63450	image/gif	44	14:57:52
21	202.139.16.45	63450	207.46.130.149	80	GET	80	14:57:52
22	207.46.130.149	80	202.139.16.45	63450	image/gif	44	14:57:52
23	202.139.16.45	63450	207.46.130.149	80	GET	74	14:57:52
24	202.139.16.45	63450	207.46.130.149	80	GET	78	14:57:53
25	202.139.16.45	63450	207.46.130.149	80	GET	83	14:57:53
26	202.139.16.45	63450	207.46.130.149	80	GET	77	14:57:53
27	207.46.130.149	80	202.139.16.45	63450	image/gif	1348	14:57:53
28	207.46.130.149	80	202.139.16.45	63450	image/gif	54	14:57:53
29	207.46.130.149	80	202.139.16.45	63450	image/gif	3966	14:57:53
30	207.46.130.149	80	202.139.16.45	63450	image/gif	1883	14:57:53
31	202.139.16.45	63450	207.46.130.149	80	GET	77	14:57:53
32	207.46.130.149	80	202.139.16.45	63450	image/gif	1044	14:57:54

FIG. 6

7/9

Site 6 Objects - Content Type "Pornography"						
Source		Destination		Type	Size	Timestamp
Object IP Address	Port	IP Address	Port			
1 202.139.16.45	63450	207.87.4.205	80	GET	36	12:01:53
2 207.87.4.205	80	202.139.16.45	63450	text/html	4254	12:01:54
3 202.139.16.45	63450	207.87.4.205	80	GET	46	12:01:56
4 207.87.4.205	80	202.139.16.45	63450	image/jpeg	4029	12:01:56
5 202.139.16.45	63450	207.87.4.205	80	GET	46	12:02:00
6 207.87.4.205	80	202.139.16.45	63450	text/html	5615	12:02:01
7 202.139.16.45	63450	207.87.4.205	80	GET	48	12:02:01
8 202.139.16.45	63450	207.87.4.205	80	GET	48	12:02:01
9 202.139.16.45	63450	207.87.4.205	80	GET	44	12:02:01
10 202.139.16.45	63450	207.87.4.205	80	GET	48	12:02:01
11 207.87.4.205	80	202.139.16.45	63450	image/gif	7026	12:02:01
12 202.139.16.45	63450	207.87.4.205	80	GET	45	12:02:02
13 207.87.4.205	80	202.139.16.45	63450	image/jpeg	20063	12:02:02
14 207.87.4.205	80	202.139.16.45	63450	image/jpeg	20526	12:02:02
15 207.87.4.205	80	202.139.16.45	63450	image/jpeg	31751	12:02:02
16 207.87.4.205	80	202.139.16.45	63450	image/gif	723	12:02:02
17 202.139.16.45	63450	207.87.4.205	80	GET	46	12:02:02
18 207.87.4.205	80	202.139.16.45	63450	image/gif	4399	12:02:03
19 202.139.16.45	63450	207.87.4.205	80	GET	46	12:02:03
20 202.139.16.45	63450	207.87.4.205	80	GET	46	12:02:04
21 207.87.4.205	80	202.139.16.45	63450	image/gif	8928	12:02:04
22 207.87.4.205	80	202.139.16.45	63450	image/gif	5474	12:02:04

FIG. 7

8 / 9

	Source		Destination				
Site 7 Objects - Content Type "Pornography"							
Object	IP Address	Port	IP Address	Port	Type	Size	Timestamp
1	202.139.16.45	63450	209.164.26.157	80	GET	38	12:02:15
2	209.164.26.157	80	202.139.16.45	63450	text/html	301	12:02:16
3	202.139.16.45	63450	209.164.26.157	80	GET	49	12:02:16
4	209.164.26.157	80	202.139.16.45	63450	text/html	4211	12:02:16
5	202.139.16.45	63450	209.164.26.157	80	GET	56	12:02:17
6	209.164.26.157	80	202.139.16.45	63450	image/gif	290839	12:02:17
7	202.139.16.45	63450	209.164.26.157	80	GET	64	12:02:21
8	209.164.26.157	80	202.139.16.45	63450	text/html	12066	12:02:21
9	202.139.16.45	63450	209.164.26.157	80	GET	76	12:02:22
10	209.164.26.157	80	202.139.16.45	63450	image/gif	4675	12:02:22
11	202.139.16.45	63450	209.164.26.157	80	GET	76	12:02:22
12	202.139.16.45	63450	209.164.26.157	80	GET	76	12:02:22
13	209.164.26.157	80	202.139.16.45	63450	image/gif	4675	12:02:22
14	209.164.26.157	80	202.139.16.45	63450	image/gif	4822	12:02:22
15	202.139.16.45	63450	209.164.26.157	80	GET	76	12:02:23
16	209.164.26.157	80	202.139.16.45	63450	image/gif	4812	12:02:23
17	202.139.16.45	63450	209.164.26.157	80	GET	76	12:02:23
18	209.164.26.157	80	202.139.16.45	63450	image/gif	4824	12:02:24
19	202.139.16.45	63450	209.164.26.157	80	GET	76	12:02:24
20	209.164.26.157	80	202.139.16.45	63450	image/gif	4858	12:02:24
21	202.139.16.45	63450	209.164.26.157	80	GET	76	12:02:24
22	209.164.26.157	80	202.139.16.45	63450	image/gif	4804	12:02:24
23	202.139.16.45	63450	209.164.26.157	80	GET	76	12:02:24
24	209.164.26.157	80	202.139.16.45	63450	image/gif	4800	12:02:24
25	202.139.16.45	63450	209.164.26.157	80	GET	76	12:02:24
26	209.164.26.157	80	202.139.16.45	63450	image/gif	4767	12:02:25
27	202.139.16.45	63450	209.164.26.157	80	GET	76	12:02:25
28	209.164.26.157	80	202.139.16.45	63450	image/gif	4732	12:02:25
29	202.139.16.45	63450	209.164.26.157	80	GET	159	12:02:25
30	202.139.16.45	63450	209.164.26.157	80	GET	159	12:02:25
31	202.139.16.45	63450	209.164.26.157	80	GET	66	12:02:26
32	209.164.26.157	80	202.139.16.45	63450	image/gif	43	12:02:26
33	202.139.16.45	63450	209.164.26.157	80	GET	66	12:02:26
34	209.164.26.157	80	202.139.16.45	63450	image/gif	93	12:02:26
35	202.139.16.45	63450	209.164.26.157	80	GET	66	12:02:26
36	209.164.26.157	80	202.139.16.45	63450	image/gif	43	12:02:26
37	202.139.16.45	63450	209.164.26.157	80	GET	66	12:02:26
38	209.164.26.157	80	202.139.16.45	63450	image/gif	7674	12:02:26
39	209.164.26.157	80	202.139.16.45	63450	image/gif	10119	12:02:26
40	202.139.16.45	63450	209.164.26.157	80	GET	76	12:02:26
41	209.164.26.157	80	202.139.16.45	63450	image/gif	4628	12:02:26
42	209.164.26.157	80	202.139.16.45	63450	image/gif	25482	12:02:26

FIG. 8

9 / 9

Model N1			
	Mean	Variance	Weight
Population	5900	18000	60
Site	Average	Std Dev	Confidence Level
1	2146	3679	45%
2	726	1974	23%
3	1703	4657	37%
4	666	1826	19%
5	1937	4741	39%
6	5149	8433	93%
7	6561	27657	87%

FIG. 9

Model N2			
Threshhold		Weight	
0.1		20	
Site	Text	Graphics	Ratio
1	20617	25938	0.79
2	6402	68319	0.09
3	25821	54455	0.47
4	1701	67037	0.03
5	19019	17252	1.10
6	9869	102919	0.10
7	19446	696989	0.03

FIG. 10

Model N3	
Threshold	Weight
30	20
Site	Word Count
1	0
2	0
3	2
4	0
5	0
6	40
7	37

FIG. 11

Site	Confidence Level
1	27%
2	34%
3	22%
4	31%
5	23%
6	96%
7	92%

FIG. 12

A.	CLASSIFICATION OF SUBJECT MATTER

Int. Cl. [7]:   G06F 17/30

According to International Patent Classification (IPC) or to both national classification and IPC

B.	FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G06F 17/30

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
WPIL, USPTO

C.	DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, A	US 5867799 A (Lang et al) 2 February 1999 Whole Document	1-13
A	US 5835905 A (Pirolli et al) 10 November 1998 Whole Document	1-13
A	US 5835722 A (Bradshaw et al) 10 November 1998	1-13

[X] Further documents are listed in the continuation of Box C	[X] See patent family annex

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance		
"E"	earlier application or patent but published on or after the international filing date	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed	"&"	document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
28 April 2000	− 9 MAY 2000

Name and mailing address of the ISA/AU	Authorized officer
AUSTRALIAN PATENT OFFICE PO BOX 200, WODEN ACT 2606, AUSTRALIA E-mail address: pct@ipaustralia.gov.au Facsimile No. (02) 6285 3929	**R.H. STOPFORD** Telephone No : (02) 6283 2177

C (Continuation).	DOCUMENTS CONSIDERED TO BE RELEVANT	
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5706507 A (Schloss) 6 January 1998 Whole Document	1-13
A	US 5678041 A (Baker et al) 14 October 1997 Whole Document	1-13

This Annex lists the known "A" publication level patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent Document Cited in Search Report							
US	5867799	US	5983214	US	6029161		
US	5835905						
US	5835722	AU	35102/97	WO	9750259		
US	5706507						
US	5678041	CA	2176775	CN	1145489	EP	748095
		JP	09/026975				

END OF ANNEX